Estimate for the number of unobserved polio infections

2022-08-22

Aim

We aim to estimated the total number of unobserved polio infections from the outbreaks in New York, given that one and only one case of paralytic polio was notified on July 18, 2022 from Rockland County. Paralysis affects a small fraction of infected individuals, with lower risk among vaccinated individuals. Therefore, even a single case of paralysis suggests substantial transmission was likely. In this short report, we attempt to back-calculate estimates (with substantial uncertainty) for the number of infections that have likely occurred by the present date.

Data

One case of paralytic polio observed in July 18, 2022. No other cases before or after this date.

Description of population

Vaccine coverage for polio is very high in Rockland county, both with OPV (in older individuals) and IPV (in younger individuals). Vaccine-induced immunity wanes with time since vaccination, and protection is against infection and paralysis are not the same from OPV and IPV.

In brief, we classify individuals as being fully susceptible, partially immune and fully immune. Partially immune individuals are susceptible to infection, but have a reduced probability of paralysis, reduced infectivity and a different shedding profile. Fully immune individuals cannot be infected at all. The proportion of individuals in each of these classes is calculated as a function of the age distribution, vaccination coverage by vaccine type, and immune waning. Thus, we have a population split into three immune categories described by proportions π_I, π_{PS}, and π_S.

A simple simulation model

The simplest model for transmission of polio, assuming oral-oral and fecal-oral transmission routes are combined into a single transmission term, is a branching process given through the discrete renewal equations. The number of new infections today is the sum of new infections caused by all infections in the past. This is given by the reproductive number at time t multiplied by the infectiousness of each currently infected person, which is a function of their time-since-infection. g gives the generation interval distribution – the probability mass function for the time between infections.

$$i'(t)^S = R_t \sum_{x<t} i(x)_S g(t-x) \tag{1}$$

This is simple to code in R:

```
tmp_infectiousness <- R0*infectiousness((t - min_t:t)-1)
inc <- sum(sapply(seq_along(min_t:t), function(x) sum(rpois(infections[x], tmp_infectiousness[x]))))
```

where `infectiousness()` is defined as a discretized gamma distribution and `infections` is a vector giving the number of individuals infected on each day prior to time t.

We could account for differences in the generation interval distribution by making g dependent on their immune state, such that partially immune individuals have a different function f. Further we can allow for R_0 to be reduced relative to the fully susceptible population by a factor ϕ:

$$i'(t)^{PS} = \phi R_t \sum_{x<t} i(x)_{PS} f(t-x) \tag{2}$$

and

$$i'(t) = i'(t)^{PS} + i'(t)^{S} \tag{3}$$

An extension allows the outbreak to be seeded with some prior population immunity. This scales the number of successful new infections by the proportion of the population which is immune.

$$i''(t)^{PS} = S(t-1)(1 - exp(-\frac{i'(t)^{PS}}{N})) \tag{4}$$

$$i''(t)^{S} = S(t-1)(1 - exp(-\frac{i'(t)^{S}}{N})) \tag{5}$$

$$i''(t) = i''(t)^{PS} + i''(t)^{S} \tag{6}$$

Again, this is straightforward to encode in R:

```
inc <- susceptible[t-1]*(1-exp(-(inc/P)))
susceptible[t] <- susceptible[t-1] - inc
```

Some proportion of infections present as cases of paralysis some time after the infections. The number of new paralysis cases observed at time t is the sum of infections that occurred x days in the past, multiplied by their probability of presenting as a case $t-x$ days later. This is determined both by the incubation period distributed $\gamma(t)$ and also the infection-paralysis-ratio (IPR), determined by parameter α. Note that the IPR is conditional on the immune state (α_{PS} and α_S).

$$y(t) = \alpha_{PS} \sum_{x<t} i''_{PS}(x)\gamma_{t-x} + \alpha_S \sum_{x<t} i''_S(x)\gamma_{t-x} \tag{7}$$

Again, this is simple to code in R. For each new infection, we simulate a probability of becoming paralytic with an incubation period as:

```
paralysis_cases <- rbinom(1, inc, prob_paralysis)
incubation_periods <- incubation_period(paralysis_cases)
```

where `incubation_period()` is defined as a discretized gamma distribution.

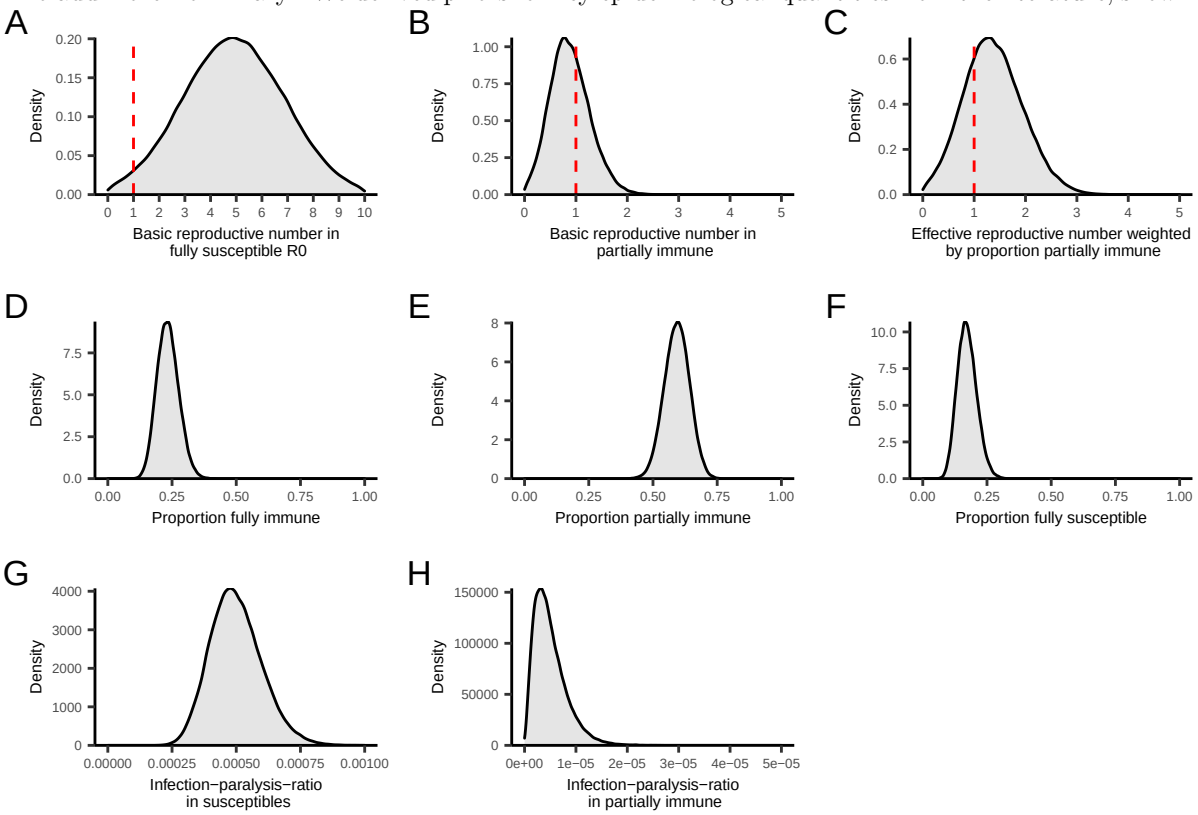
Inference using the simple simulation model

Fitting this stochastic model to the observed data (which is a trivial dataset) gives an estimate for the total number of polio infections. An accurate estimate requires reliable estimates, or at least priors, for the generation interval distribution, incubation period distribution, the infection-paralysis-ratio, the proportion of the population in each vaccination class and their immunity, and how all of these parameters vary across immune classes.

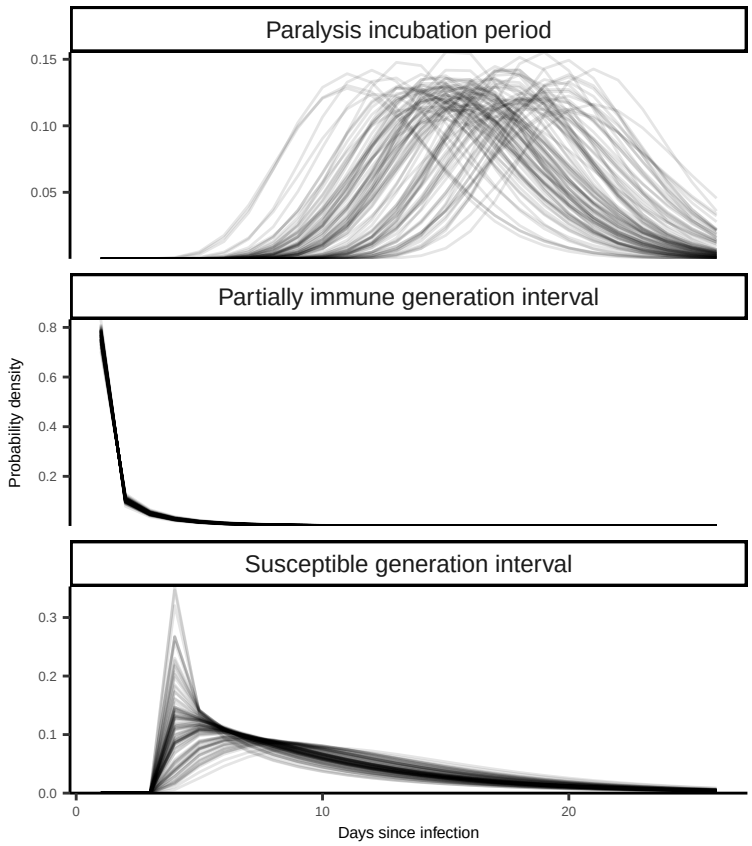
A first-pass fitting approach is to use an approximate Bayesian computation approach. We draw parameter values from prior distributions for each of these parameters, store draws which generate trajectories that are entirely consistent with the data (one case of paralysis followed by weeks of no cases), and discard trajectories which are inconsistent. The simulations are stochastic, and thus some trajectories die out with $R_e < 1$, yet still generate a case of paralysis. The result gives an approximation for the posterior distribution of each parameter conditional on the observed data.

Priors

« To add more from Mary» We derived priors for key epidemiological quantities from the literature, shown here:

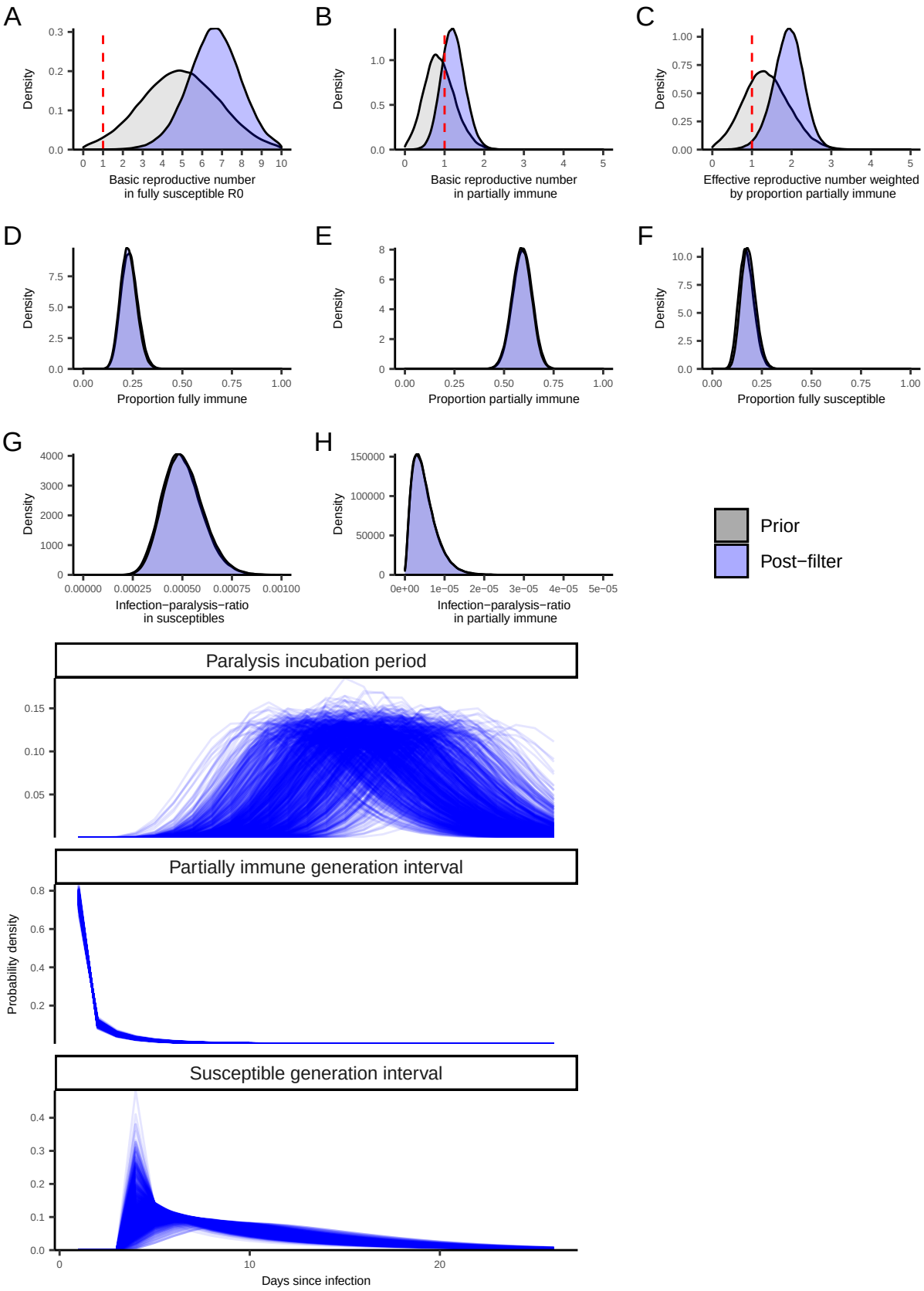


And for the generation interval and incubation period distributions:



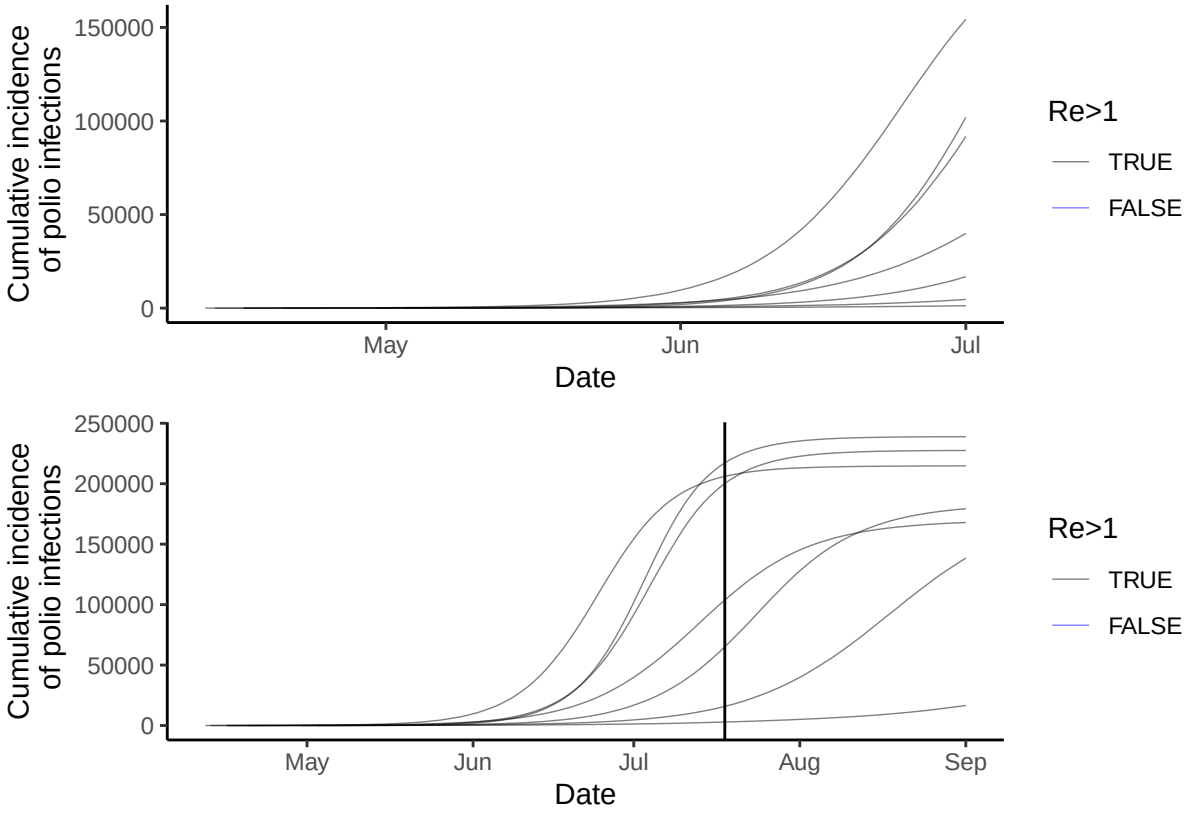
Post fitting

We find that only trajectories with a very high R_0 are consistent with the data, as the high level of population immunity renders the effective reproductive number very low.

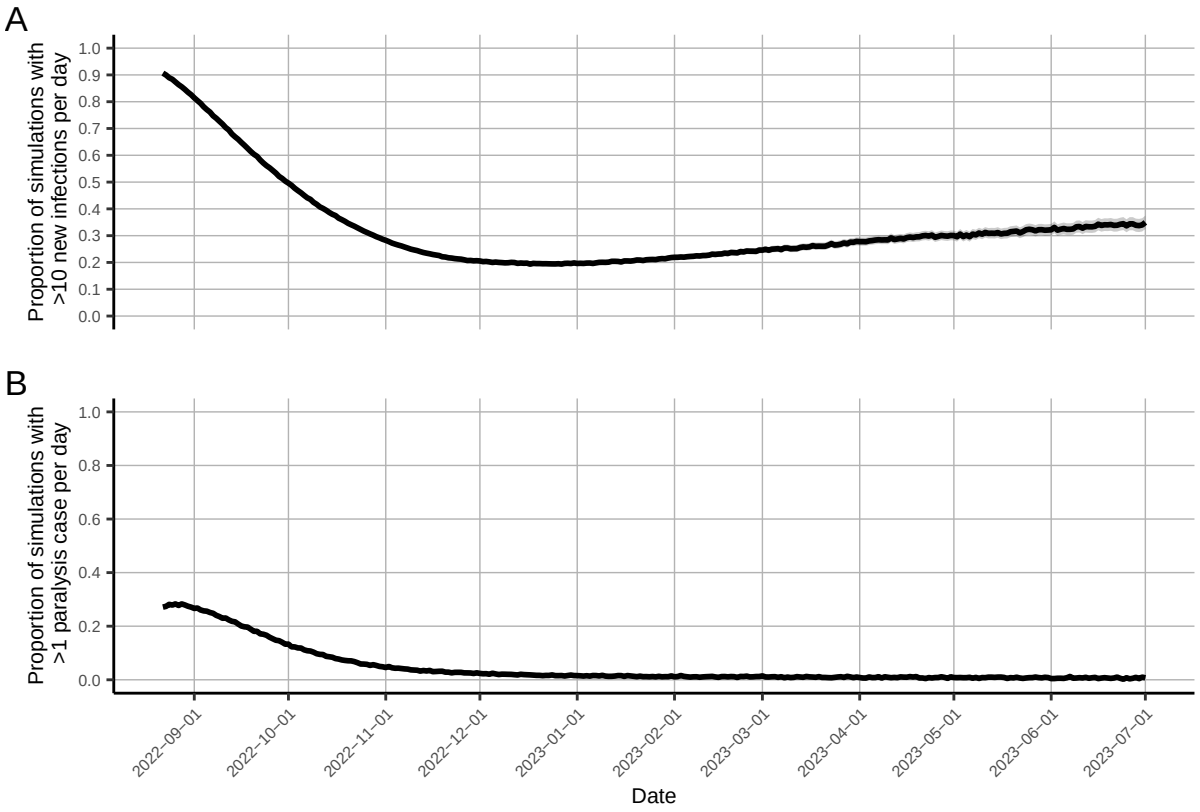


A subset of possible trajectories demonstrates just how variable the trajectory could have been to have

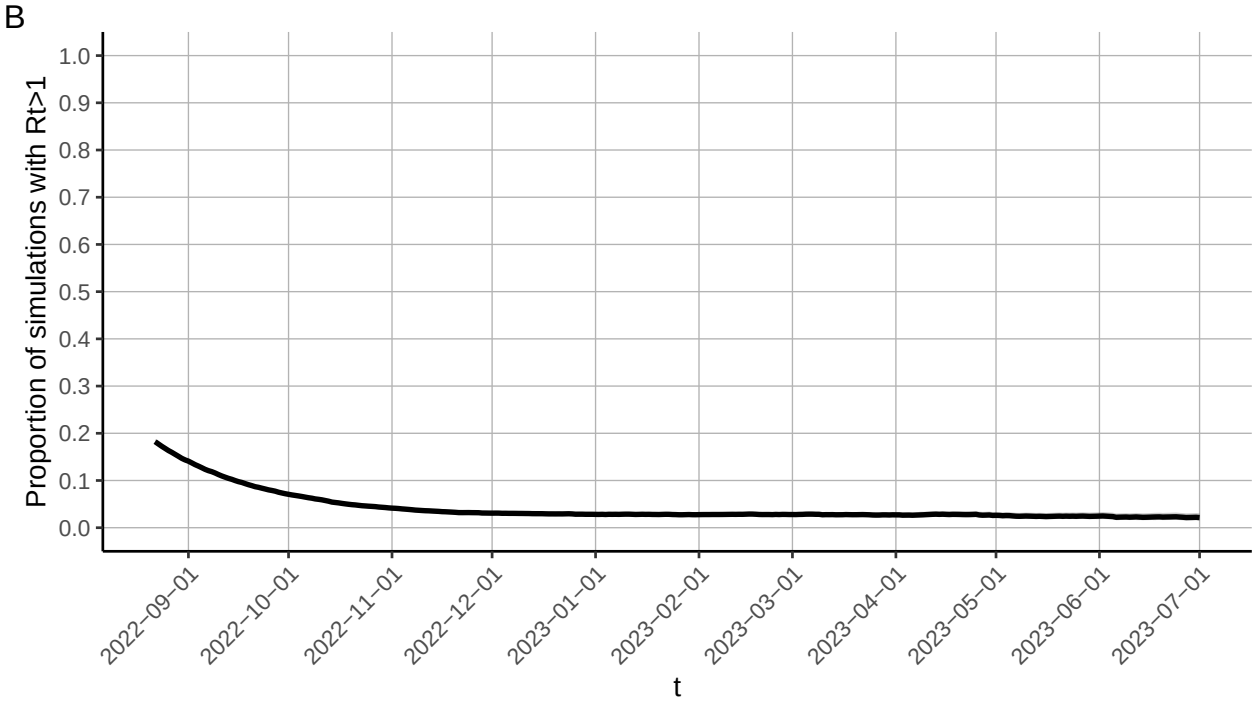
generated one case of paralysis.



An easier visualization is to inspect the proportion of runs consistent with the data which are still generating new cases over time.



And Rt.



The proportion of trajectories with $R_t > 1$ by 2022-08-20:

```
## [1] 0.01517737
```

And with incidence >0 by 2022-08-20:

```
## [1] 0.07646543
```

Number of infections by 2022-08-20.

Mean infections	Median infections	Lower 95% PrI	Lower 80% PrI	Upper 80% PrI	Upper 95% PrI
161215.2	191646.5	772.8	20508.9	228894.2	241554.9

Number of paralysis cases by the end of the simulation (ie. once the epidemic has burned through the susceptible population).

Mean paralysis cases	Median paralysis cases	Lower 95% PrI	Lower 80% PrI	Upper 80% PrI	Upper 95% PrI
24.2	24	6	13	36	45

Proportion of simulations with more than one case of paralysis after the initial case.

Vaccine strategy	Proportion
1	0.984

Number of infections by the end of the simulation (ie. once the epidemic has burned through the susceptible population).

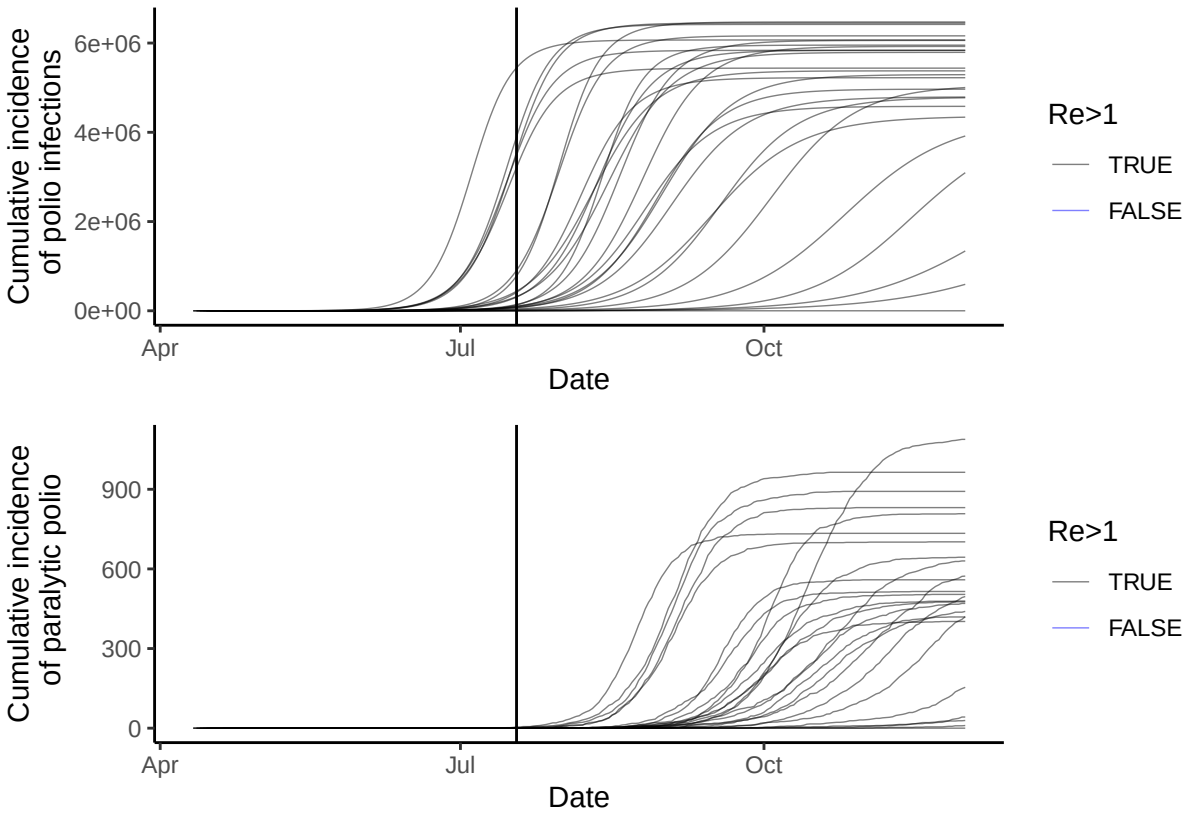
Mean infections	Median infections	Lower 95% PrI	Lower 80% PrI	Upper 80% PrI	Upper 95% PrI
190518.9	200247.5	62873.3	143033.2	230103.8	242259

Proportion of simulations with more than 100 further infections.

Vaccine strategy	Proportion
1	0.994

Dynamics in New York City

To provide a comparator scenario where the polio epidemic continues to grow and is seeded with sustained transmission in nearby counties (as is suggested by wastewater data), we ran the simulation using the same posterior draws, but for the population size and immune class distribution of New York City.


Cumulative incidence of polio infections
6e+06
4e+06
2e+06
0e+00
Apr
Jul
Oct
Date
Re>1
TRUE
FALSE
Cumulative incidence of paralytic polio
900
600
300
0
Apr
Jul
Oct
Date
Re>1
TRUE
FALSE